Final Project - STA457

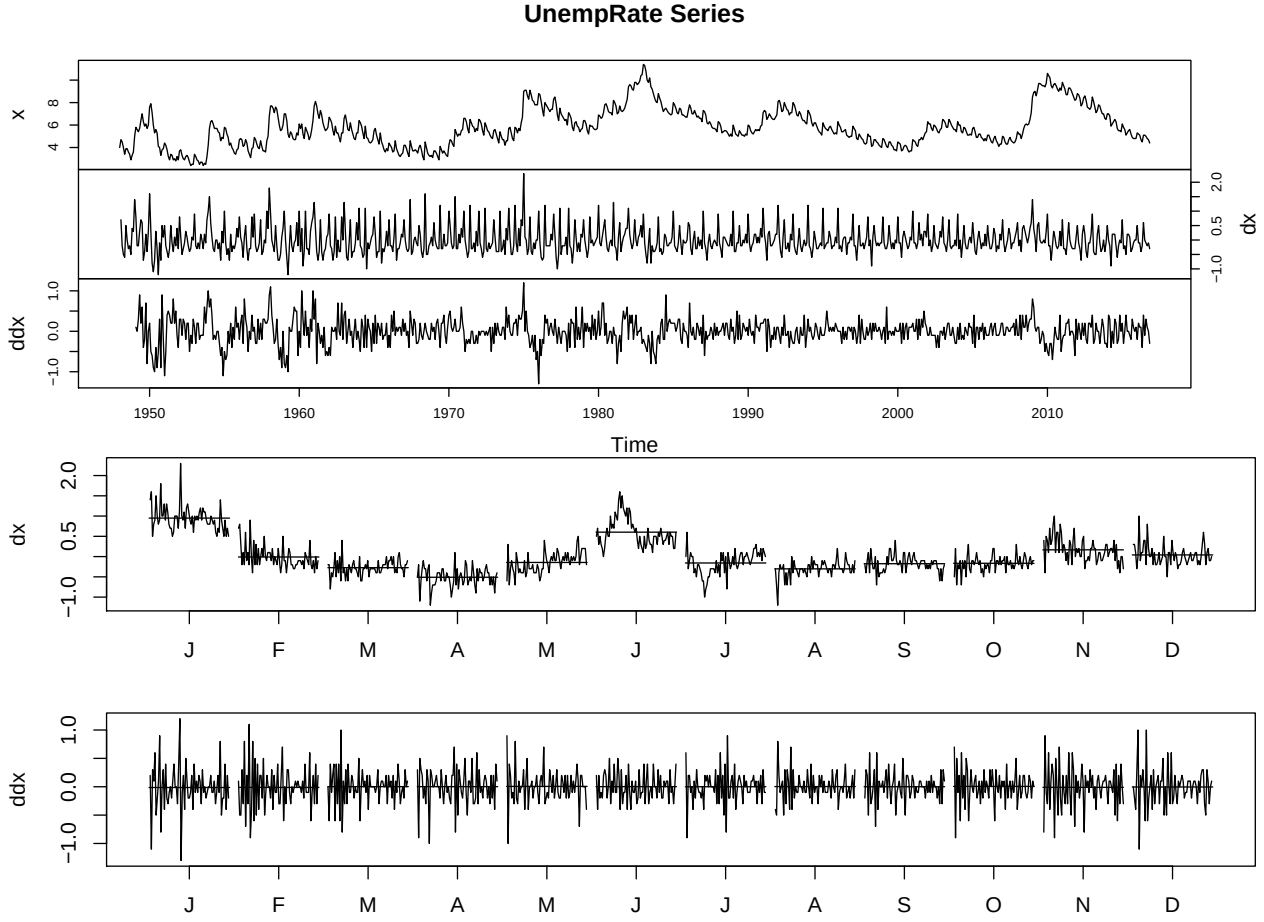
Luis Rojas

17/12/2020

1.) For this project we use the time series *UnempRate*. This Series is the monthly unemployment rate for the U.S. from January 1948 to November 2016 for a total of 827 observations.

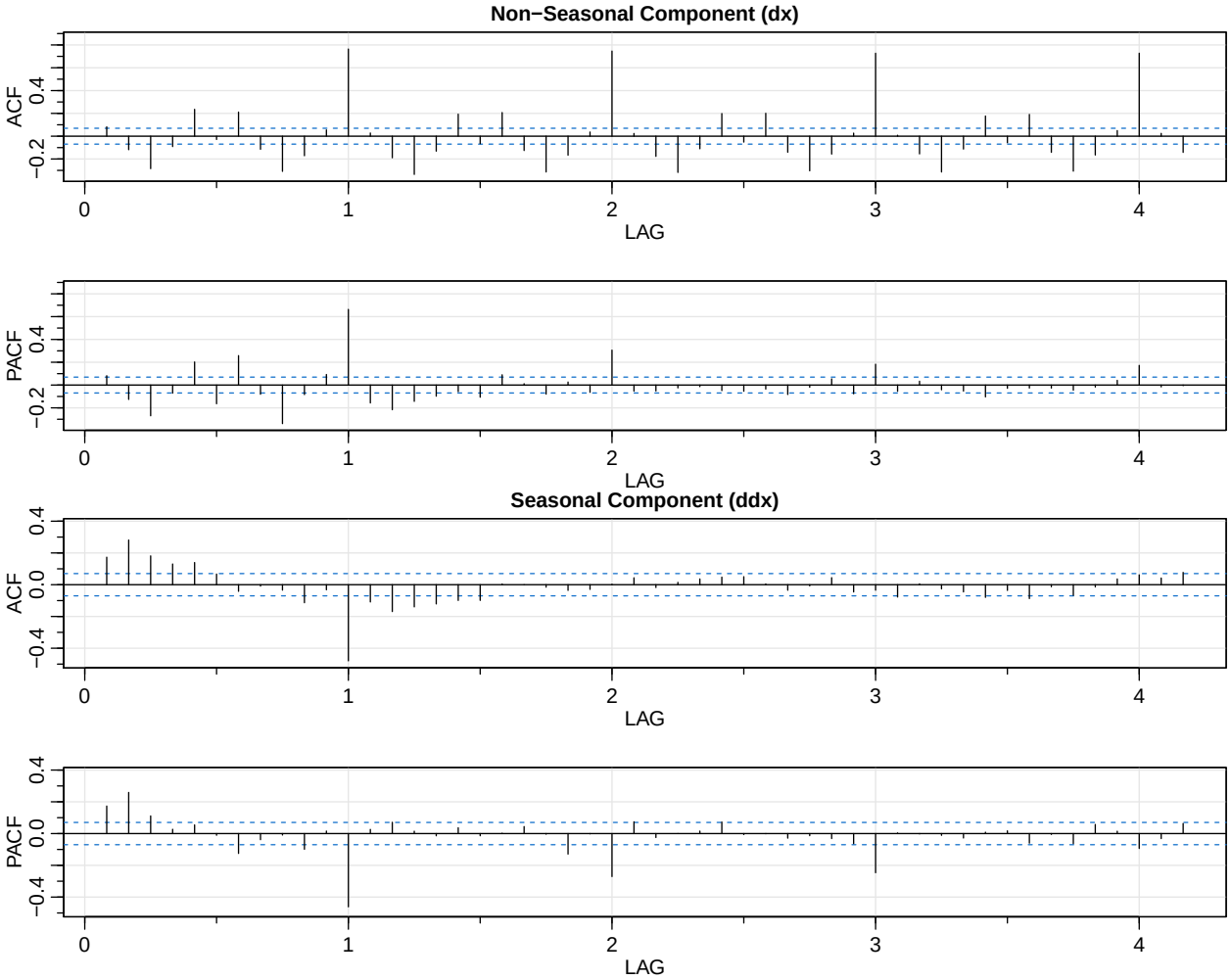
2.) We first plot the data to analyze the observations, we note that the series has an slightly upward trend.

3.) Then, the differenced series ("dx") removes the trend making the series stationary. Note that since the series is given in rates, so this difference makes the series a growth rate series. The series comes from data collected in a monthly basis, therefore we can appreciate in the second set of plots that in dx there is an approximate persistence in the seasons ($\mathbf{dx_t \approx dx_{t-12}}$). Then, the twelfth-order difference is applied (ddx), from this we can see that a seasonal ARIMA model is appropriate in this case since the series becomes stationary.



4.) Since the model presents a seasonal component we analyze the seasonal and non-seasonal ACF and PACF plots for both.

The seasonal component shows a ACF plot that appears to cuts off after lag $1s$ ($s = 12$) and the PACF appears to be tailing off at lags $1s$, $2s$, $3s$,... . Therefore, we decide for a **SMA(1)** model for the seasonal component that is $\mathbf{P = 0}$, $\mathbf{Q = 1}$, in the season (s=12).



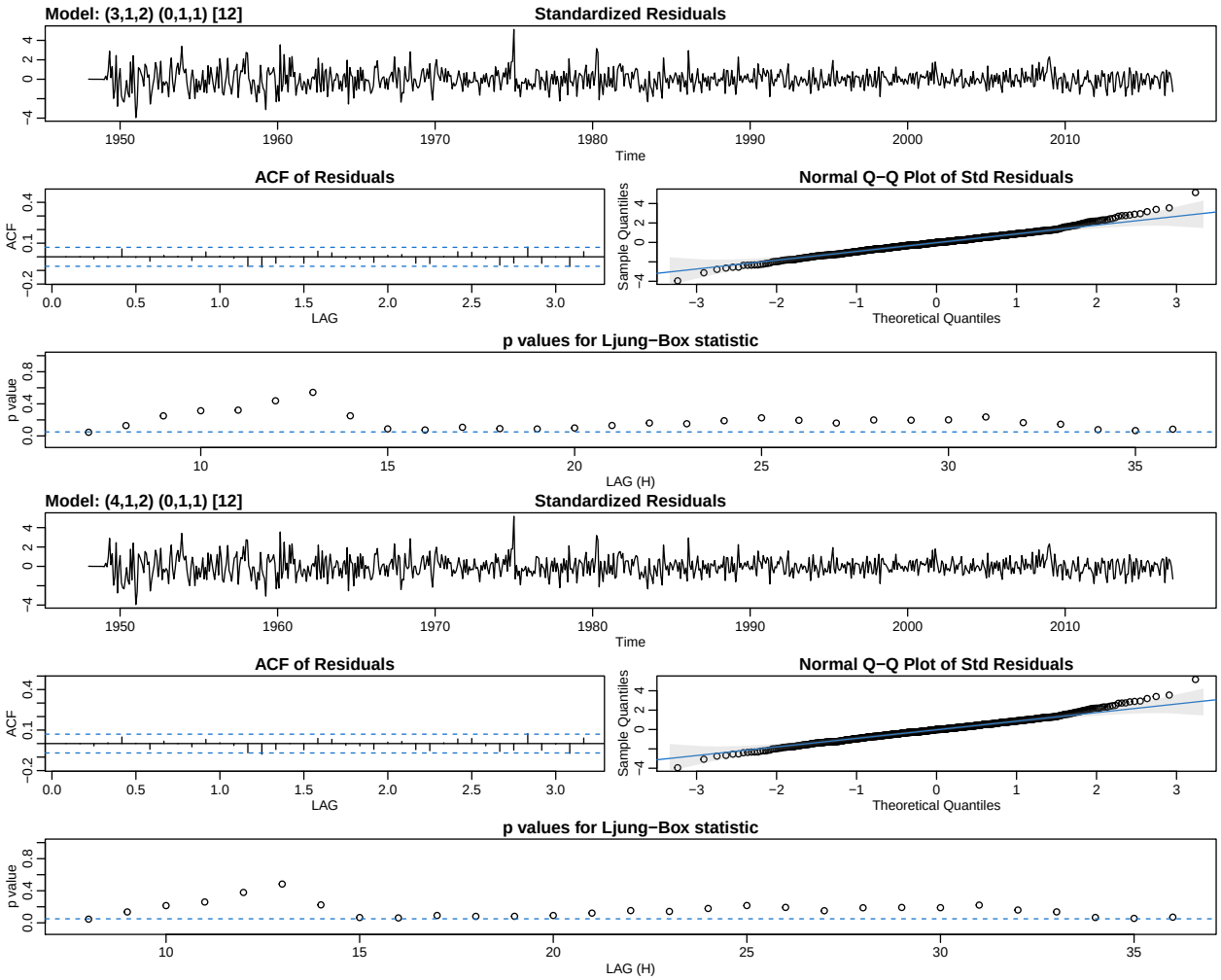
For the non-seasonal component is more difficult to decide for a model based only in the ACF and PACF plots due to the nature of the data. For this reason instead of deciding for a model and then analyze the parameters we implement inverse engineering setting a number of model and deciding in the best model using the lowest AIC.

After running the corresponding for loop we obtain the table following table of AIC and choose the model with the lowest AIC.

	q = 0	q = 1	q = 2	q = 3	q = 4	Smallest AIC
p = 0	1094.593	1089.7043	1083.2028	1005.7389	1006.7903	1005.7389
p = 1	1091.024	1091.0191	1065.3062	1007.4241	978.8931	978.8931
p = 2	1079.787	1052.3267	806.8970	943.8449	943.0731	806.8970
p = 3	1019.307	1019.5479	799.5341	938.3247	940.2164	799.5341
p = 4	1017.070	978.6816	788.8394	940.1259	940.8258	788.8394

Then we decide between the models with the two lowest AIC for the non-seasonal component, we choose the **ARIMA(3,1,2)** and **ARIMA(4,1,2)** for the non-seasonal component.

5.) - 6.) - 7.) Then for the proposed models we use the SARIMA function to determine for the most appropiate model. The first model $\mathbf{ARIMA(3,1,2)x(0,1,1)_{12}}$ appears to produce good estimates but since there are clearly some missing values we cannot determine the the significance of the parameters, then it is discarded immediately.

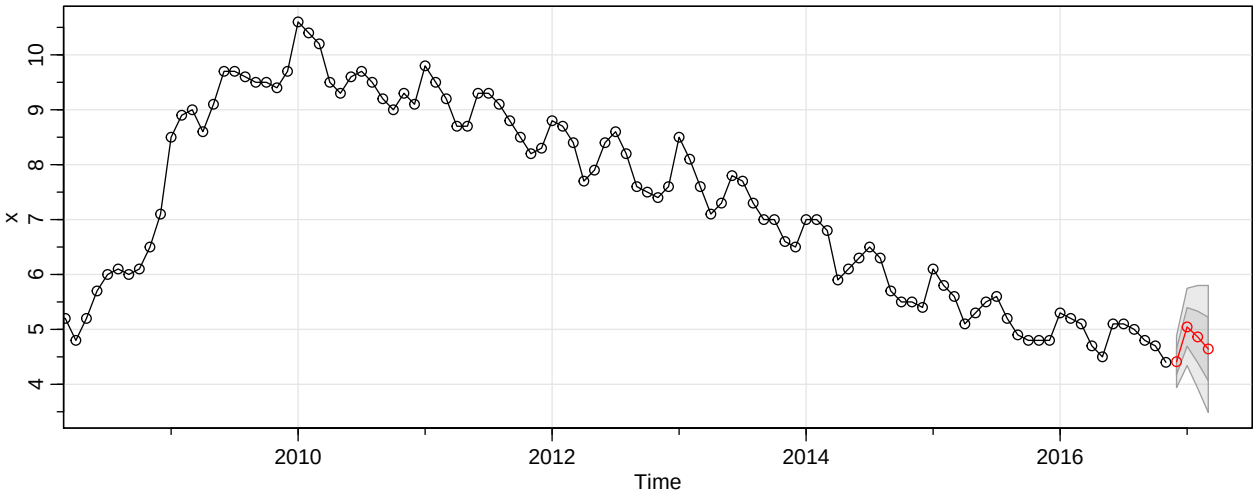


For the second model $\mathbf{ARIMA(4,1,2)x(0,1,1)_{12}}$ the inspection of the standardized residuals plot reveals that there are very few observations falling outside the 3 standard deviations in magnitude and there appears to be no obvious patterns. The ACF plot shows that the model does not have autocorrelation in the errors, therefore no departure from the model assumptions. The normal Q-Q plot of the residuals shows that the normality assumption in the error term is reasonable. Finally, the plot for the p-values for the Ljung-Box Q-test shows that the p-values are never significant at the lags show failing to rejecting the null hypothesis that the errors are independently distributed showing no autocorrelation among the errors, so no departure from the model assumptions. The model appears to fit well therefore all the information criteria and diagnosis plots prefer the $\mathbf{ARIMA(4,1,2)x(0,1,1)_{12}}$ model.

Therefore, the model $\mathbf{ARIMA(4,1,2)x(0,1,1)_{12}}$ yields the following coefficients:

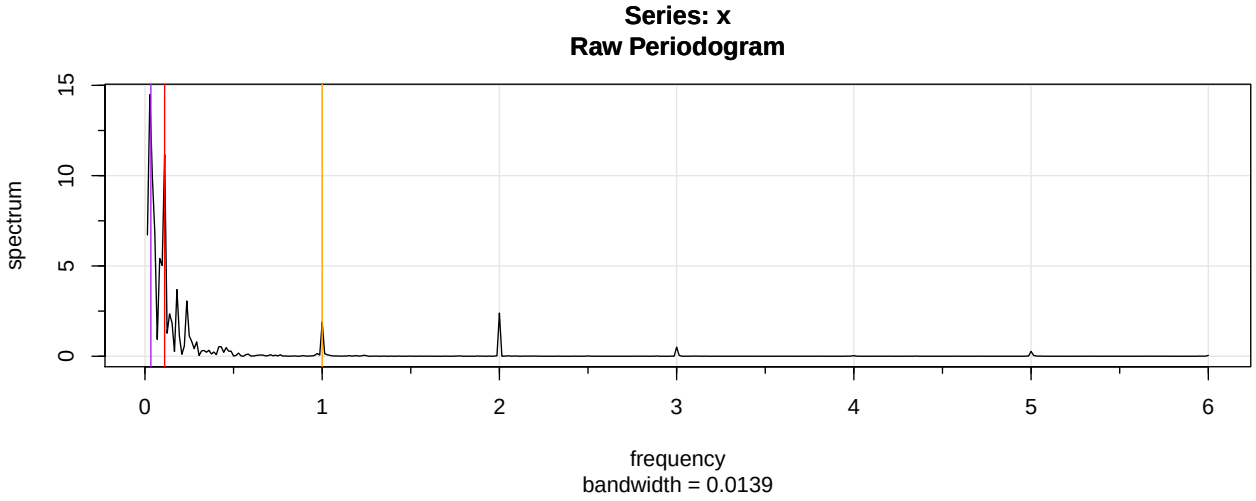
	Estimate	SE	t.value	p.value
ar1	-0.2243	0.2215	-1.0127	0.3115
ar2	0.7052	0.1503	4.6914	0.0000
ar3	0.0998	0.0739	1.3501	0.1774
ar4	-0.0337	0.0586	-0.5747	0.5657
ma1	0.3327	0.2186	1.5224	0.1283
ma2	-0.4798	0.1575	-3.0463	0.0024
sma1	-0.7646	0.0262	-29.1403	0.0000

8.) Then, we forecast the next 4 periods. Here we present a plot and the corresponding table containing the predicted values with the corresponding confidence intervals.



Prediction_ahead	Predicted_Values	Lower_Bound	Upper_Bound
1	4.411210	3.949252	4.873168
2	5.045021	4.355372	5.734670
3	4.863015	3.945640	5.780389
4	4.643506	3.510124	5.776889

9.) We identify the dominant frequencies by sprectral analysis and present the 95% confidence intervals for the first three dominant frequencies.



The series *UnempRate* is taken in a monthly basis, so the frequency is 12, the Δ is the reciprocal of the frequency that is $\Delta = \frac{1}{12}$. The observations between January 1948 to November 2016 are n = 827, however FFT transform uses a number of redundancies in the calculation of the DFT, since this n is not a factor of 2, we pad the detrended data of length 459 to the next highly composite integer n' by adding zeros, then the new n is $n' = 864$. The dominant frequency for this series according to the periodogram is located at $\frac{1}{30}$, followed by other two lower frequencies at $\frac{1}{9}$ and 1.

Note that the frequency axis is labeled in multiples of $\Delta = \frac{1}{12}$, therefore the omegas are given by $\omega_{\mathbf{1}} = \frac{\mathbf{1}}{\mathbf{12}} * \frac{\mathbf{1}}{\mathbf{30}} = \frac{\mathbf{1}}{\mathbf{360}}$ or one cycle every 30 years for the 864 observations in the series, in this case the periodogram at this frequency is equivalent to $\omega_{\mathbf{1}} = \frac{\mathbf{1}}{\mathbf{360}} = \frac{\frac{\mathbf{12}}{\mathbf{5}}}{\mathbf{864}}$ note that for security we use specification at $\frac{12}{5} \approx 2.4$.

Similarly, $\omega_{\mathbf{2}} = \frac{\mathbf{1}}{\mathbf{12}} * \frac{\mathbf{1}}{\mathbf{9}} = \frac{\mathbf{1}}{\mathbf{108}}$ or one cycle every 9 years, that is scaled for 864 observations the omega is equivalent to $\omega_{\mathbf{2}} = \frac{\mathbf{1}}{\mathbf{108}} = \frac{\mathbf{8}}{\mathbf{864}}$. The third frequency correspond to $\omega_{\mathbf{2}} = \frac{\mathbf{1}}{\mathbf{12}} * \frac{\mathbf{1}}{\mathbf{9}} = \frac{\mathbf{1}}{\mathbf{108}}$ or a yearly cycle, if we scaled this specification by the 864 observations and obtain $\omega_{\mathbf{3}} = \frac{\mathbf{1}}{\mathbf{12}} = \frac{\mathbf{72}}{\mathbf{864}}$. The confidence interval at the 95% level for the cycle of 30 years or the spectrum $f_S(\frac{1}{30})$ is [3.93, 572.04] and for the spectrum at the 9 year cycle, $f_S(\frac{1}{9})$, [3.03, 441.27]. Finally the confidence interval for the yearly cycle, $f_S(\frac{1}{1^2})$, [0.51, 74.65]. While these confidence intervals are quite large to be of much use, we can still argue that the three peaks are significant, since the zero is not contained in the confidence intervals.

Omega	Lower_Bound	Upper_Bound
0.0027778	3.9261016	572.04461
0.0092593	3.0285584	441.26991
0.0833333	0.5123411	74.64961

10.) A brief discussion of the results

This unemployment rate time series for the USA becomes stabilized when we apply the first difference and the to control for the seasonal component we apply the twelfth-order difference. After analzying the ACF and PACF plots and also a set of model for the non-seasonal component we decide for $\mathbf{ARIMA}(\mathbf{4,1,2})\mathbf{x}(\mathbf{0,1,1})_{\mathbf{12}}$ model. This model meets the assumptions pretty well and fits the data accordingly to our expectations. The next four forecast share the common patter of the series and its confidence interval is not too wide so the model seem to be doing a good job. The model presents patterns that are better reflected when using a periodogram to analyize the frequencies, finding relevant frequencies at the 30, 9 and 1 year cycles, all of them significant.

Appendix: All R-code for this report.

```
knitr::opts_chunk$set(echo = F)
knitr::opts_chunk$set(fig.width=10, fig.height=4)
options(warn=-1)
library(astsa)
library(kableExtra)

#Questions 1 - 2 - 3

x    = UnempRate
dx   = diff(x)
ddx  = diff(dx, 12)
plot.ts(cbind(x, dx, ddx), main="UnempRate Series", ylim = c(-20, 10),  yax.flip = TRUE)

# below of interest for showing seasonal random walk:
par(mfrow=c(2,1), mar = c(2.5,4,1.5,1))
monthplot(dx)
monthplot(ddx)

#Question 4

par(mfrow=c(2,2), mar = c(2.5,4,1.5,1))
#Non-Seasonal Component
a <- capture.output(acf2(dx, 50, main = "Non-Seasonal Component (dx)"))
#Seasonal Component
b <- capture.output(acf2(ddx, 50, main = "Seasonal Component (ddx)"))


#Question 5 - 6 - 7
#Create a table of AIC values to decide for the best model

AICtable1 <- matrix( NA, 5, 6 )
dimnames( AICtable1 ) <- list( c( paste( "p =", 0 : 4 ) ),
                               c( paste ( "q =", 0 : 4 ), "Smallest AIC") )
for ( p in 0:4 ) {
  for( q in 0:4 ){
    AICtable1[p+1, q+1] <- AIC( arima( diff(x), order = c(p,0,q) ) )
    }
  AICtable1[p+1, 6] <- min( AICtable1[p+1,1:5] )
}
knitr::kable(AICtable1)


M <- capture.output(sarima(x, 3,1,2, 0,1,1, 12,details = TRUE))  # model 1
N <- capture.output(lr <- sarima(x, 4,1,2, 0,1,1, 12)) #Best Model
knitr::kable(lr$ttable)


#Question 8

prediction <- sarima.for(x, 4, 4,1,2, 0,1,1,12)  # forecasts
```

```r
#Lower bound confidence interval
Lower_Bound <- prediction$pred - 1.96 * prediction$se

#Upper bound confidence interval
Upper_Bound <- prediction$pred + 1.96 * prediction$se

#Making a nice table to present C.I. for each prediction
Predicted_Values <- prediction$pred
Prediction_ahead <- c(1:4)
knitr::kable(cbind(Prediction_ahead, Predicted_Values, Lower_Bound, Upper_Bound))

# Question 9

delta <- 1/frequency(x) #1/12
n <- nextn(length(x)) #n = 827, n' = 864

x.per = mvspec(x, log="no")

abline(v= 1/30, lty="solid", col = "purple")
abline(v= 1/9, lty="solid", col = "red")
abline(v= 1, lty="solid", col = "orange")

omega_1 <- (1/30) * delta #1/360
omega_2 <- (1/9) * delta #1/108
omega_3 <- (1) * delta #1/12

U = qchisq(.025,2)    # 0.05063
L = qchisq(.975,2)    # 7.37775
Lower_Bound <- c(2*x.per$spec[omega_1 * n]/L, 2*x.per$spec[omega_2 * n]/L, 2*x.per$spec[omega_3 * n]/L)

Upper_Bound <- c(2*x.per$spec[omega_1 * n]/U, 2*x.per$spec[omega_2 * n]/U, 2*x.per$spec[omega_3 * n]/U)

#the confidence interval then is:
Omega <- c(omega_1, omega_2, omega_3)
knitr::kable(cbind(Omega, Lower_Bound, Upper_Bound))
```